

Evaluating the Effectiveness of Military Decision
Support Systems: Theoretical Foundations,
Expert System Design, and
Experimental Plan

Antonio Leal

Integrated Sciences Corporation

BASIC RESEARCH



U. S. Army

Research Institute for the Behavioral and Social Sciences

September 1982

83 09 29 037

Approved for public release; distribution unlimited.

This report, as submitted by the contractor, has been cleared for release to Defense Technical Information Center (DTIC) to comply with regulatory requirements. It has been given no primary distribution other than to DTIC and will be available only through DTIC or other reference services such as the National Technical Information Service (NTIS). The views, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other official documentation.

AD-A133080
DTIC FILE COPY

DTIC
ELECTED
SEP 30 1983
A

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Research Note 83-18	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) EVALUATING THE EFFECTIVENESS OF MILITARY DECISION SUPPORT SYSTEMS: THEORETICAL FOUNDATIONS, EXPERT SYSTEM DESIGN, AND EXPERIMENTAL PLAN		5. TYPE OF REPORT & PERIOD COVERED --
7. AUTHOR(s) Antonio Leal (Integrated Sciences Corporation)		6. PERFORMING ORG. REPORT NUMBER 345-3
9. PERFORMING ORGANIZATION NAME AND ADDRESS Integrated Sciences Corporation 1640 Fifth Street Santa Monica, CA 90401		8. CONTRACT OR GRANT NUMBER(s) MDA903-81-C-0449
11. CONTROLLING OFFICE NAME AND ADDRESS U.S. Army Research Institute for the Behavioral and Social Sciences 5001 Eisenhower Avenue, Alexandria, VA 22333		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 2Q161102B74F
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) --		12. REPORT DATE September 1982
		13. NUMBER OF PAGES 65
		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE --
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report) --		
18. SUPPLEMENTARY NOTES The contracting officer's technical representative was Dr. Stanley M. Halpin		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Expert systems Training aids Game Planning aids Simulation		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) The main objective of this program is to construct a flexible test bed for the evaluation of the effectiveness of computer-based expert systems in military training and planning. The technical approach consists of simulating the characteristics of expert systems in a game-like environment. Such characteristics include (1) friendly and English-like (stylized) dialogue, (2) system explanations of rationale about decision recommendations, (3) an ability to make (Continued)		

relevant suggestions and comments about situation assessments and about plans proposed by the user, and (4) the use of high level strategic concepts and terminology.

The required software for such a program includes (1) a game environment simulator, (2) a simulated expert system for the game, and (3) an evaluation program for recording execution histories and for summarization. The game simulator will contain provisions for the experimenter to adjust critical parameters so that a controlled environment can be maintained. The expert system will monitor the progress of the game and can be interrogated as the user sees fit. A facility will also be provided for evaluating the user's performance under different modes of consultation with the expert system.

[illegible]

TABLE OF CONTENTS

I. INTRODUCTION	1
1.1 Program Objective	1
1.2 Approach	1
1.3 Progress	2
II. THEORETICAL FOUNDATIONS	3
2.1 Expert Systems with Ground Truth	3
2.2 Scenario Generator Requirements	4
2.3 The General Model	5
III. THE EXPERT SYSTEM	9
3.1 Optimal Decision Tables	9
3.2 Expert Rule Formation	13
3.3 Expert Rules for the Simulation Game	16
IV. EXPERIMENTAL PLAN	25
4.1 Objective	25
4.2 Verifiable Hypotheses	25
4.3 Test Factors	27
4.4 Performance Measures	29
4.5 Test Design	29
4.6 Experimental Procedure	30
APPENDIX A. Decision Diagram for Game	A-1
APPENDIX B. Decision Tables for Game	B-1

LIST OF FIGURES

Figure 1. Decision Diagram Elements	7
Figure 2. Table 1 Decisions	10
Figure 3. Player Decisions	12
Figure 4. Decision Table Clusters	15
Figure 5. Example Clustering	17
Figure 6. Decision Situations: First Rule Number	19
Figure 7. Expert System Game Rules	21
Figure 8. Potential Benefit of Expert System	26
Figure 9. Latin Square Design	31
Figure 10. Contingency Plans for one-factor 2-Level Experiments	32
Figure 11. One-Way Analysis of Variance	33

I. INTRODUCTION

1.1 Program Objective

→ The main objective of this program is to construct a flexible testbed for the evaluation of the effectiveness of computer-based expert systems in military training and planning.

1.2 Approach

The technical approach consists of simulating the characteristics of expert systems in a game-like environment. Such characteristics include (1) friendly system user interaction, (2) system explanations of rationale about decision recommendations, (3) an ability to make relevant suggestions and comments about situation assessments and about plans proposed by the user, and (4) the use of high-level strategic concepts and terminology.

The required software for such a program includes (1) a game environment simulator called the "Scenario Generator", (2) a simulated expert system for the game called the "Expert Aid", (3) an Optimality Algorithm for computing the best decisions in any situation, and (4) an Evaluation Module for recording execution histories and performance parameters. The expert system will monitor the progress of the game and can be interrogated as the user sees fit. A facility will also be provided for evaluating the user's performance under different modes of consultation with the expert system.

↑

1.3 Progress

To date, the following progress has been made on the design and implementation of the expert system.

- (1) Design of the Scenario Generator.
- (2) Prototype implementation of the Scenario Generator game for testing purposes.
- (3) Completion of the Optimality Algorithm Design.
- (4) Implementation and execution of Optimality Algorithm resulting in a set of optimal decision tables.
- (5) Design of the Expert System and integration into the existing Scenario Generator.
- (6) Design of the Experimental Plan.

Plans for project continuation include the following major tasks.

- (1) Implementation of the full expert system including system explanations of recommendations.
- (2) Implementation and integration of the Evaluation Module for measuring user performance.
- (3) Completion of the evaluation experiments.

II. THEORETICAL FOUNDATIONS

2.1 Expert Systems with Ground Truth

In a problem environment involving multiple decision points, "ground truth" is a complete description of the optimal strategy for all possible situations the decision maker may encounter. These optimal decisions must be identifiable by computation and must not be dependent upon the judgment of humans regardless of their expertise. An expert system, armed with ground truth, could make perfect recommendations to the decision maker at all times. If the environment contains uncertain events, the ground truth can be accurate only up to the highest expected utility. That is, ground truth may not guarantee successful outcomes; it just means that no other decision strategy could produce better performance in the long run.

Ground truth, then, is a property of the environment and not of the expert system. Most expert systems do not have built-in ground truth. Their recommendations are based on human knowledge and experience encoded as rules. The reason for this is obvious. Any environment complex enough to warrant the construction of an expert system is usually too complex to compute ground truth. However, the existence of ground truth permits the calibration of performance measures in expert system evaluation. Learning curves for new users could be generated by comparing decision performance with and without the recommendations of the expert system, but the results would be unreliable since the user's expertise continuously increases simply from contact with the environment. On the other hand, ground truth provides an upper bound on performance and a standard against which performance can

be measured. In an environment containing uncertain events, evaluation without ground truth must rely on the outcomes of the decisions rather than on the quality of the decisions themselves. The results would then be clouded by the random processes within the environment. With ground truth as a standard, precise measurements of decision performance can be attained regardless of the outcomes of the experiments.

2.2 Scenario Generator Requirements

The task, then, is to devise a decision environment which is complex enough to be challenging to a decision maker and, yet, simple enough to permit the calculation of ground truth. These objectives lead to the following requirements for the expert system testbed Scenario Generator.

- (1) Environment complex enough to be challenging to a decision maker.
- (2) Environment simple enough to calculate ground truth.
- (3) Environment amenable to the construction of and use of an expert system for decision recommendations.
- (4) Environment containing uncertain events.
- (5) Environment containing information feedback during the decision process.
- (6) Environment within the military domain and which embodies a military decision problem.

Since uncertain events and information feedback are common elements of decision environments, their inclusion in the Scenario Generator is important for the transfer of evaluation results to real-world expert systems.

The selected environment that satisfies the above criteria is a military decision game. An expert system designed around a game would provide recommendations about the best "moves" in different situations. Ground truth can be established (if the game is not too complex) since the criteria for winning and losing are specified precisely in the rules. The only disadvantage is that games are not generally very realistic. However, as will be shown in the next section, a methodology can be described which will permit the construction of a ground truth expert system in any environment. The experimental results will then be transferrable.

2.3 The General Model

The scenario game environment is founded upon a general model for a class of ground truth expert systems. The major components of the model that characterize the environment are:

- (1) Well-defined decision points with all choice options specified in advance.
- (2) Well-defined event points with all outcomes and associated probabilities specified.

- (3) Well-defined information feedback points with all possible combinations of "reports" specified.

Figure 1 shows notational symbols for each of the components: decision points, event points, and report points. An "identification point" is a place during the decision process where an important observation is made when it does not coincide with one of the other components. By connecting symbols together to form a tree, a class of decision environments can be constructed which leads directly to the design of the ground truth expert system.

A complete decision structure should have the following characteristics.

- (1) Start with a decision point.
- (2) Possess normalized probabilities on each branch leading out from an event point.
- (3) Possess utility values on every termination branch.

Once built, the decision structure dictates the procedure for calculating the ground truth. At every decision point, the value of the decision (for all possible situations) is equal to the maximum of the values on its branches. At every event point, the value of the event is equal to the expected value of the utilities and probabilities on its branches. At every report point, the utility of the information is equal to the utility of its branch given the probability of the report. Thus, decision utilities are calculated by a "rollback" procedure that begins at the terminal branches and ends at the initial decision point. These calculations must be performed for every combination of decision situation and every combination of report outcome. In other words, a single

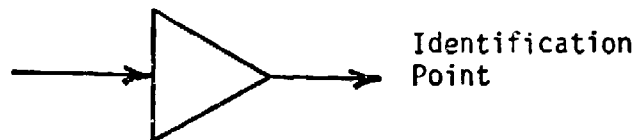
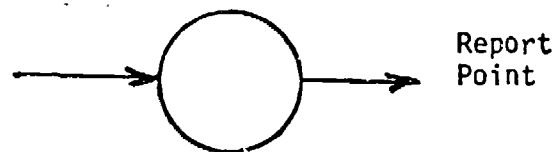
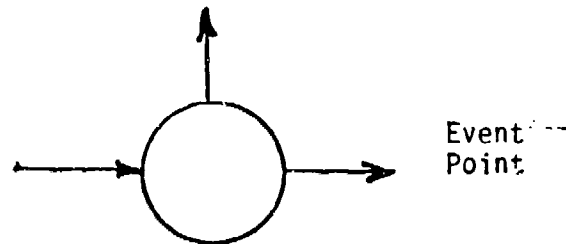
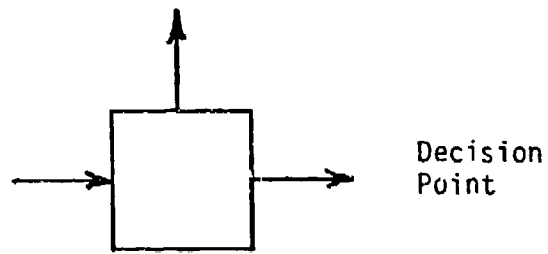


Figure 1. Decision Diagram Elements

tree diagram shows the sequential decisions that must be made, but the decision maker could be faced with the sequence over and over again in a repetitive environment in which situational parameters (such as resources, etc.) are different each time. This leads to a complex series of matrices necessary to rollback the utility values and calculate the optimal decisions at every decision point. The "Military Breakthrough" game, designed as an experimental testbed, satisfies the requirements of the general model and illustrates how these utility calculations are made and how they lead to the construction of the corresponding expert system.

III. THE EXPERT SYSTEM

3.1 Optimal Decision Tables

The 140 decision tables provide the optimal decisions for any situation encountered in the simulation game. Appendix B contains a complete listing of all tables arranged on 28 pages: one page for each combination of enemy line (1-4) and enemy encampment (1-7). Five tables are associated with each situation.

Table 1: Attack/Continue/Shallow/Deep

Table 2: Attack/Continue given Shallow, not detected

Table 3: Attack/Continue given Shallow, detected

Table 4: Attack/Continue given Deep, not detected

Table 5: Attack/Continue given Deep, detected

Table 1 is used when the player encounters a new enemy encampment and is faced with the decision of whether or not to send out reconnaissance (See Figure 2). If reconnaissance is chosen, a secondary decision is required to determine the type: shallow (S) or deep (D). However, if the player decides not to send reconnaissance, he is left with the decision to immediately attack (A) or continue (C) to the next enemy position. The optimal decisions for these 4 choices appear in Table 1.

The actual optimal decision depends upon 3 factors: (1) the current position, (2) the current strength of the friendly forces, and (3) the first report (X1) received by the player on the status of the strength of the enemy forces. Thus, decision Table 1 on each of the 28 pages in the listing in the Appendix is 12x64 in size, one row for each of the 12 possible friendly force combinations and one column for each of the 64 possible X1 reports. The last line, showing a friendly force of zero, is omitted since the game would be over and no further decisions could be made.

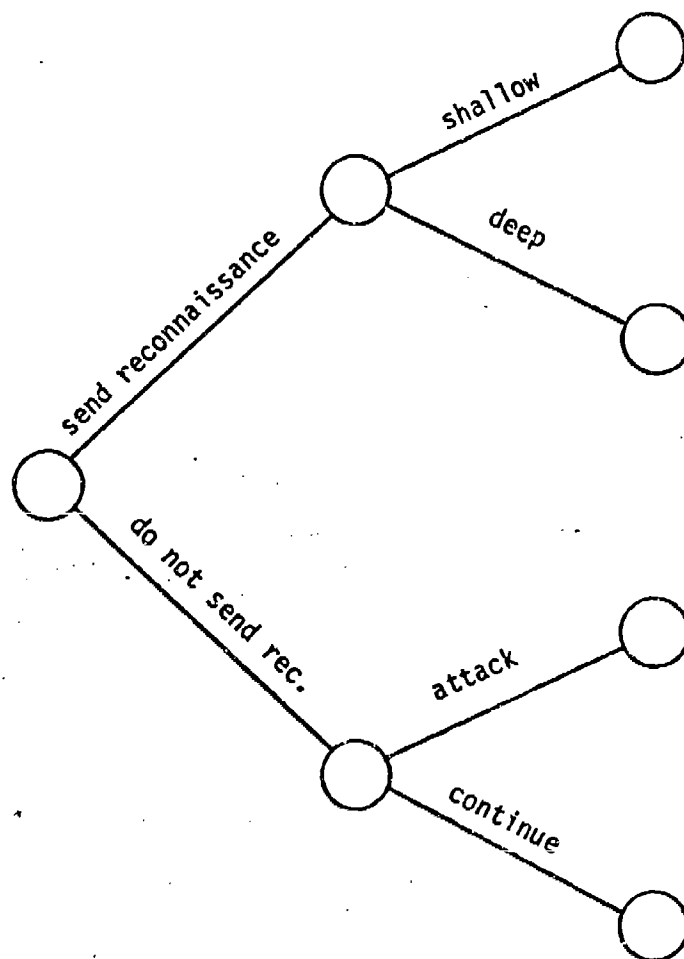


Figure 2. Table 1 Decisions

For example, if the player is facing enemy line #1 and enemy encampment #3 with a force of MHH (medium strength artillery, high strength infantry, and high strength tanks) and observes that the report for the enemy strength is NNL (nil, nil, low reading from top-to-bottom), the optimal decision would be "D", ("take Deep reconnaissance").

Tables 2 and 3 are used only when the player has chosen shallow (S) reconnaissance for the initial decision (Table 1). Table 2 is used if the reconnaissance force is not detected and Table 3 is used if the force is detected. For these two tables, only the decisions of attacking (A) or continuing (C) are relevant. The column labels at the top of the page still apply, but now represent the outcome of report X2, a more accurate portrayal of the true enemy strength. Some of the entries in Tables 2 and 3 are missing since these particular combinations are impossible for report X2.

Tables 4 and 5 at the bottom of each page are used if deep (D) reconnaissance is made for the initial decision in Table 1. Again, two tables are required, Table 4 for "not detected" and Table 5 for "detected". In these cases, however, only 4 columns are necessary since the true enemy strength is known to the player at this time (as the result of report X3) and can be only 1 of 4 possible combinations.

Thus, as can be seen in Figure 3, S and D are intermediate decision points whereas A and C are final decision points. Numbers inside circles refer to the Tables used.

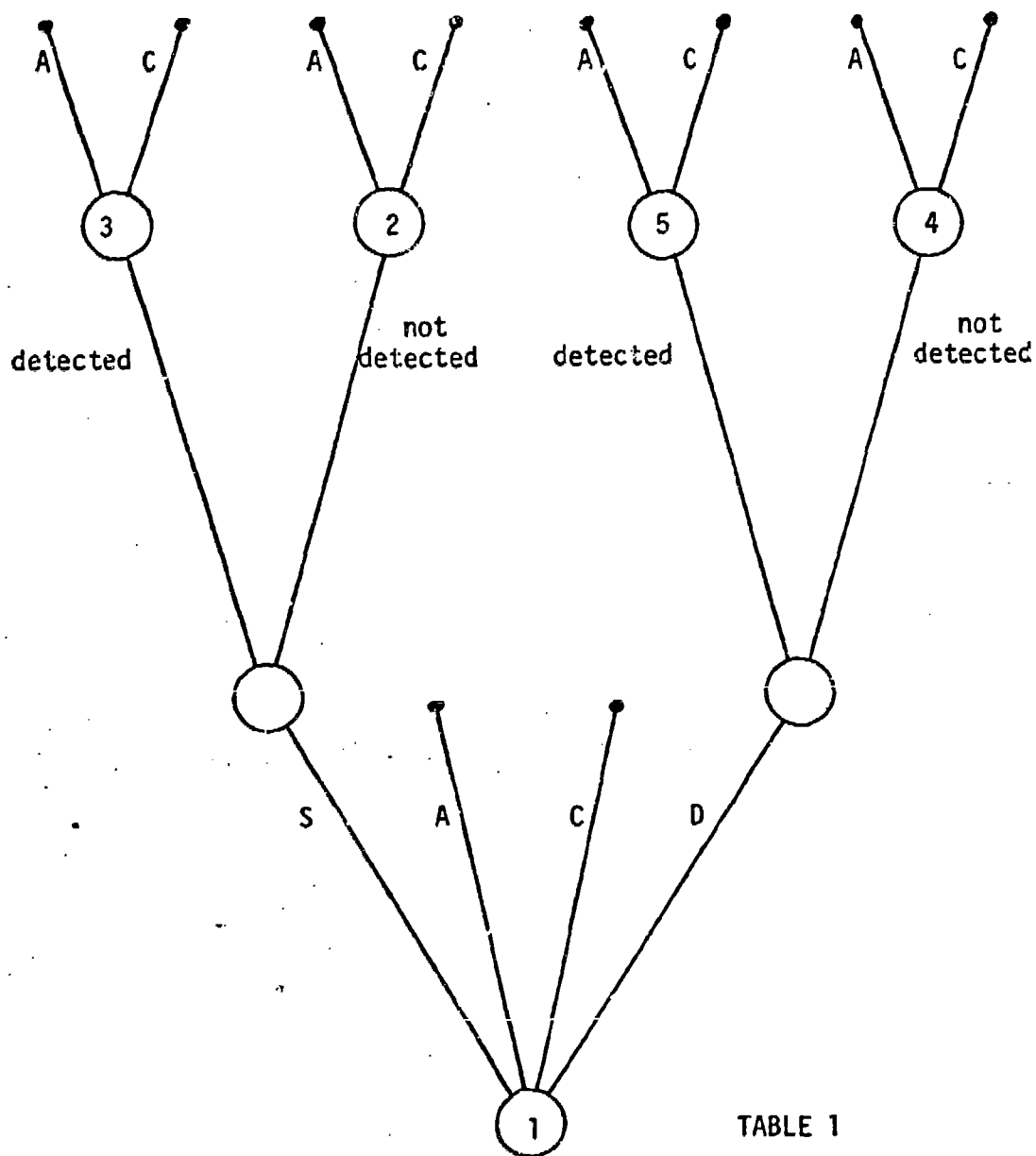


TABLE 1

Figure 3. Player Decisions

3.2 Expert Rule Formation

An expert system, which had access to all 140 decision tables, could give optimal recommendations for every encountered situation. However, such information may not be the most valuable to a user. One of the purposes of an expert system is to teach the human user enough about the characteristics of the application environment so that he acquires some expertise. As they now exist, the decision tables are far too complex to learn, especially by point recommendations as individual situations are encountered during simulation exercises.

A more useful set of recommendations would be based on general rules that capture most, but not necessarily all of the optimal decisions. By observing the optimal tables listed in Appendix B, it is clear that most of the decisions fall into clusters. Thus, an expert system "rule" can be defined as a single recommended decision which represents the majority of optimal decisions within a specified region of the decision space.

The accuracy of the expert system will ultimately depend upon the uniformity of decisions within selected clusters. The accuracy can, thus, be calculated precisely. It is simply the total percentage of correctly classified decision points. The clusters should be chosen in such a way to produce high accuracy while reducing the cognitive complexity associated with learning the rules.

3.3 Expert Rules for the Simulation Game

The decisions for a single table are composed of a number of connected rules, one rule for each rectangular region. Figure 5 shows the divisions for encampment #3 in enemy line #3. Table 1 has 3 distinct rectangular clusters and 1 non-rectangular region covering the remainder of the table. To specify the rules, only the boundaries for the columns and rows are necessary. Thus, the upper left-hand cluster (with mostly A's) falls between 0 and 2 in enemy strength and between 7 and 9 in friendly strength. A set of 3 rules would be required to completely describe Table 1. The following form shows this specification using "E" for the enemy strength and "F" for the friendly strength.

<u>RULE #</u>	<u>≤ E ≤</u>	<u>≤ F ≤</u>	<u>DO</u>	<u>ELSE</u>
1	0 2	7 9	A	2
2	0 3	5 5	A	3
3	0 9	0 4	S	D

The above description can be read as follows:

- (1) If the enemy strength is between 0 and 2 and the friendly strength is between 7 and 9, attack (A), otherwise follow Rule 2.
- (2) If the enemy strength is between 0 and 3 and the friendly strength is equal to 5, attack, otherwise follow Rule 3.
- (3) If the enemy strength is between 0 and 9 and the friendly strength is between 0 and 4, send shallow reconnaissance (S), otherwise send deep reconnaissance (D).

The last decision recommendation in the last rule (D) covers the non-rectangular portion of the table.

Obviously, the accuracy of the resulting rules depends on the accuracy of the individual regions. For the example table described above, the accuracy of the rules is shown in the following chart.

<u>Region</u>	<u>Correct</u>	<u>Total</u>	<u>% Correct</u>	<u>Decision</u>
1	41	50	82%	A
2	36	40	90%	A
3	79	128	62%	S
4	484	506	96%	D
Total	640	704	91%	

Thus, the expert rules for this particular decision table are 91% correct.

The figures on the following pages list all 115 rules contained in the current preliminary version of the Expert System. Figure 6 catalogues the 5 main tables for each of the 28 positions and their associated first rule numbers. For example, if the player is facing enemy line #3 and encampment #1 and is considering a decision about reconnaissance (Table 1), the applicable rule is #68.

Figure 7 contains the actual rules. If both enemy strength conditions (E) and both friendly strength conditions (F) are true, the decision under the "DO" column should be followed where "A" means attack, "S" means take shallow reconnaissance, "D" means take deep reconnaissance, and "C" means continue (no attack). If any of the 4 conditions fail, the instructions under the "ELSE" column should be followed. These instructions are either (1) a decision: A, S, D, or C, or (2) a reference number which indicates the next rule to follow.

<u>Enemy Line</u>	<u>Enemy Encampment</u>	<u>Table</u>	<u>Rule #</u>	<u>Enemy Line</u>	<u>Enemy Encampment</u>	<u>Table</u>	<u>Rule #</u>
1	1	1	1	2	1	1	35
1	1	2	2	2	1	2	40
1	1	3	2	2	1	3	41
1	1	4	3	2	1	4	42
1	1	5	3	2	1	5	3
1	2	1	1	2	2	1	43
1	2	2	4	2	2	2	40
1	2	3	2	2	2	3	45
1	2	4	3	2	2	4	42
1	2	5	3	2	2	5	3
1	3	1	5	2	3	1	46
1	3	2	4	2	3	2	47
1	3	3	6	2	3	3	45
1	3	4	3	2	3	4	42
1	3	5	3	2	3	5	42
1	4	1	7	2	4	1	48
1	4	2	10	2	4	2	51
1	4	3	4	2	4	3	54
1	4	4	11	2	4	4	23
1	4	5	12	2	4	5	12
1	5	1	13	2	5	1	56
1	5	2	19	2	5	2	60
1	5	3	21	2	5	3	62
1	5	4	23	2	5	4	23
1	5	5	12	2	5	5	12
1	6	1	24	2	6	1	64
1	6	2	31	2	6	2	67
1	6	3	33	2	6	3	61
1	6	4	34	2	6	4	34
1	6	5	23	2	6	5	23

Figure 6. Decision Situations: First Rule Number

<u>Enemy Line</u>	<u>Enemy Encampment</u>	<u>Table</u>	<u>Rule #</u>	<u>Enemy Line</u>	<u>Enemy Encampment</u>	<u>Table</u>	<u>Rule #</u>
3	1	1	68	4	1	1	96
3	1	2	71	4	1	2	99
3	1	3	41	4	1	3	45
3	1	4	12	4	1	4	11
3	1	5	3	4	1	5	12
3	2	1	72	4	2	1	100
3	2	2	74	4	2	2	105
3	2	3	45	4	2	3	45
3	2	4	12	4	2	4	11
3	2	5	12	4	2	5	12
3	3	1	75	4	3	1	106
3	3	2	76	4	3	2	20
3	3	3	4	4	3	3	45
3	3	4	12	4	3	4	23
3	3	5	12	4	3	5	12
3	4	1	78	4	4	1	108
3	4	2	81	4	4	2	20
3	4	3	45	4	4	3	74
3	4	4	23	4	4	4	23
3	4	5	12	4	4	5	11
3	5	1	83	4	5	1	109
3	5	2	61	4	5	2	32
3	5	3	88	4	5	3	20
3	5	4	23	4	5	4	34
3	5	5	12	4	5	5	23
3	6	1	89	4	6	1	111
3	6	2	94	4	6	2	113
3	6	3	61	4	6	3	114
3	6	4	34	4	6	4	34
3	6	5	23	4	6	5	34

Figure 6. (Continued)

<u>Rule #</u>	<u>≤</u>	<u>E</u>	<u>≤</u>	<u>≤</u>	<u>F</u>	<u>≤</u>	<u>DO</u>	<u>ELSE</u>
1	0		1	9		9	S	D
2	0		3	9		9	A	C
3	0		1	6		9	A	C
4	0		3	6		9	A	C
5	0		1	9		9	A	D
6	0		3	8		9	A	C
7	0		2	8		9	A	8
8	0		2	5		5	A	9
9	3		9	5		5	S	D
10	0		3	4		9	A	C
11	0		1	0		9	A	C
12	0		1	4		9	A	C
13	0		2	9		9	A	14
14	0		4	8		8	A	15
15	0		2	7		7	A	16
16	0		4	4		5	A	17
17	5		9	4		5	S	18
18	0		9	0		2	S	D
19	3		4	6		7	C	20
20	5		9	0		9	C	A
21	0		4	9		9	A	22
22	0		2	4		8	A	C
23	0		3	0		9	A	C
24	0		5	7		9	A	25
25	0		1	6		6	A	26
26	0		5	4		5	A	27
27	0		3	0		2	A	28
28	2		5	6		6	D	29
29	6		6	6		9	D	30
30	8		9	0		5	D	S

Figure 7. Expert System Game Rules

<u>Rule #</u>	<u>≤</u>	<u>E</u>	<u>≤</u>	<u>≤</u>	<u>F</u>	<u>≤</u>	<u>DO</u>	<u>ELSE</u>
61	5		6	8		8	A	87
62	0		2	0		5	A	63
63	3		4	0		7	C	20
64	6		9	6		9	D	65
65	4		5	6		6	D	66
66	5		9	0		5	S	A
67	5		6	6		6	C	32
68	0		1	9		9	S	69
69	0		2	5		5	A	70
70	0		9	0		4	S	D
71	3		3	4		8	C	52
72	0		0	7		9	A	73
73	0		3	5		5	A	70
74	0		2	4		9	A	C
75	0		2	7		9	A	73
76	3		4	6		7	C	77
77	0		4	0		4	C	20
78	0		3	5		9	A	79
79	4		4	5		5	A	80
80	0		9	4		4	S	D
81	3		4	6		6	C	82
82	0		4	0		2	C	20
83	4		4	6		6	D	84
84	0		3	0		4	A	85
85	4		9	0		4	S	86
86	5		9	5		9	D	A
87	0		4	0		9	A	C
88	0		2	0		2	C	63
89	7		8	9		9	S	90
90	5		7	0		2	S	91

Figure 7. (Continued)

<u>Rule #</u>	<u>≤</u>	<u>E</u>	<u>≤</u>	<u>≤</u>	<u>F</u>	<u>≤</u>	<u>DO</u>	<u>ELSE</u>
31	3		6	6		6	C	52
32	7		9	0		9	C	A
33	3		4	6		6	C	20
34	0		6	0		9	A	C
35	0		1	9		9	S	36
36	0		1	5		5	A	37
37	2		3	5		5	S	38
38	0		3	0		4	D	39
39	4		9	0		5	S	D
40	0		2	5		9	A	C
41	0		2	7		9	A	C
42	0		1	5		9	A	C
43	0		0	9		9	A	44
44	1		1	9		9	S	36
45	0		2	6		9	A	C
46	0		1	8		9	A	8
47	0		3	5		9	A	C
48	0		3	7		9	A	49
49	0		3	5		5	A	50
50	4		5	5		5	S	D
51	3		3	6		7	C	52
52	0		3	0		2	C	53
53	4		9	0		9	C	A
54	3		3	6		8	C	55
55	0		3	0		5	C	53
56	0		3	7		9	A	57
57	4		5	8		8	A	58
58	0		3	0		5	A	59
59	4		5	0		5	S	D
60	3		4	6		6	C	61

Figure 7. (Continued)

<u>Rule #</u>	<u>≤</u>	<u>E</u>	<u>≤</u>	<u>≤</u>	<u>F</u>	<u>≤</u>	<u>DO</u>	<u>ELSE</u>
91	6		7	4		5	S	92
92	8		9	0		5	D	93
93	6		9	6		8	D	A
94	8		9	0		9	C	95
95	6		7	0		4	C	A
96	0		4	0		2	S	97
97	0		3	4		5	A	98
98	8		9	6		7	S	A
99	0		2	0		9	A	C
100	2		9	0		2	S	101
101	0		1	0		2	A	102
102	0		3	4		5	A	103
103	0		1	6		9	A	104
104	8		9	6		8	S	D
105	3		3	6		8	C	53
106	4		9	4		9	D	107
107	3		9	0		2	S	A
108	5		9	4		9	D	107
109	6		9	4		9	D	110
110	4		9	0		2	S	A
111	5		9	0		2	S	112
112	6		9	4		8	D	A
113	7		9	0		8	C	A
114	7		9	0		9	C	115
115	5		6	0		5	C	A

Figure 7. (Continued)

IV. EXPERIMENTAL PLAN

4.1 Objective

The objective of the experimental plan is to formulate a controlled experimental environment and procedure for assessing the expert system as a training aid and as a performance aid. The experiments are designed to verify a number of hypotheses about the usefulness of the expert system in various modes of operation. Human subjects will be enlisted to participate in the experiments and an automatic performance measurement module will be installed into the expert system to monitor the decisions made by the subjects. The following sections describe the experimental plan in more detail.

4.2 Verifiable Hypotheses

Four verifiable hypotheses have been selected for experimentation.

- H₁: Performance improves with increased level of aiding.
- H₂: As task familiarity increases, the usefulness of the expert system as a training aid increases.
- H₃: When expertise is high, the benefit of the expert system as a performance aid is lower.
- H₄: The maximum level of expertise is reached sooner when using the expert system.

Figure 8 shows the hypothesized relationship between the user's learning curve (dotted line) and the benefit of an expert system (solid line). When the user is new to the environment, aiding

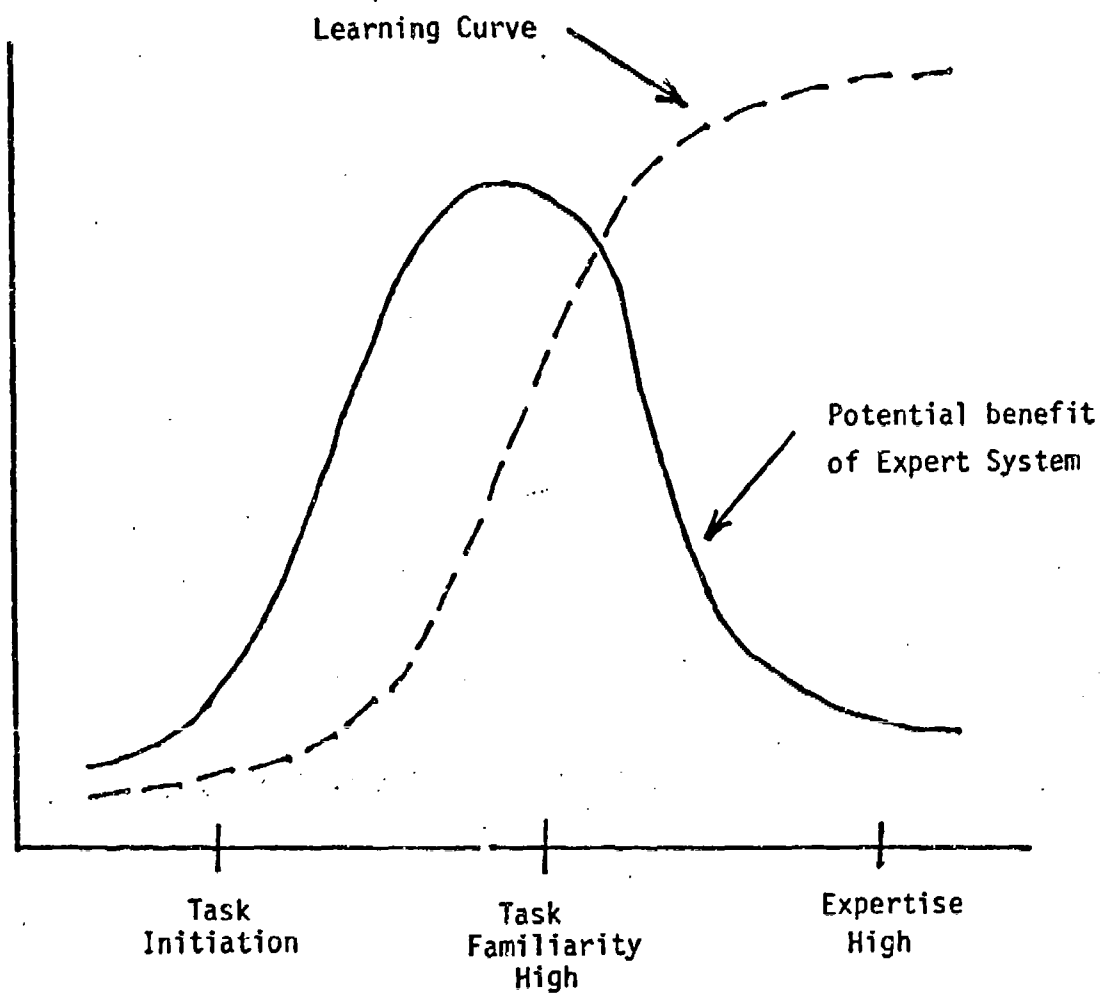


Figure 8. Potential Benefit of Expert System

of a particular subject. The subject simply plays one or more games with no aiding at all and his expertise level is the average number of correct decisions made divided by the total number decisions made.

$$\text{expertise} = \frac{\text{correct decisions}}{\text{total decisions}}$$

A second measure of expertise is based on the actual outcomes of the decision. This measure is simply the subject's average game score over a number of successive trials.

$$\text{expertise} = \text{average game score}$$

"Task Familiarity" can be measured by combining the number of games played with the total amount of time interacting with the game environment. The "Maximum Level of Expertise" is reached when the increase in expertise from successive trials approaches zero.

$$\text{task familiarity} = f(\# \text{ games, total time})$$

max expertise reached when:

$$\text{expertise}_i - \text{expertise}_{i-1} < \delta$$

The amount of aiding chosen is simply the number of times the subject used the expert system divided by the number of times he/she could have used it (in those experiments in which there is a choice).

$$\text{amount of aiding} = \frac{\# \text{ times E.S. used}}{\# \text{ times E.S. available}}$$

Preceding Page Blank

4.4 Performance Measures

Measures of level of performance correspond to or are functions of the measured test factors. These can be simple or aggregated measures.

- (1) Decision time as a function of expertise
- (2) Decision time as a function of task familiarity
- (3) Learning rate (increase in task familiarity over time)
- (4) Mastery rate (increase in expertise over time)
- (5) Time to reach maximum level of expertise
- (6) Increased game score over time
- (7) Expertise vs. amount of aiding used

4.5 Test Design

The test design specifies those factors that will be the objects of the test and those factors that will be eliminated in order to remove their influence on the results.

For hypothesis H_1 (See Section 4.2), the effects of factor 1 on factor 3 (See Section 4.3) will be measured. Factors 2, 4, 5, and 6 will be eliminated. Three levels of expert system aiding will be employed: full, partial, and none. With full aiding, the subjects will be able to consult the expert system any time desired. With partial aiding, only selected decisions will permit aiding. For experiments concerning H_1 , a Latin Square-type design will be used.

For hypotheses H_2 and H_3 , the effects of factors 1 and 6 will be assessed on factor 3. This means that the rate of change in level of expertise will be measured. Factors 2, 4, and 5 will be eliminated. Repeating single factor experiments will be used.

For hypothesis H_4 , the effects of factor 1 on factor 4 will be measured. Factors 2, 3, 5, and 6 will be eliminated. Again, single factor experiments with repetitions will be used.

4.6 Experimental Procedure

Every experiment will be carried out in 4 phases: (1) definition, (2) design, (3) data collection, and (4) data analysis. The definition phase will concern the precise plan for carrying out each experiment. This includes the dependent and independent variables (factors), the number of observations taken, the number of subjects, the number of experimental groups, etc. The design phase concerns the type of test performed: Latin Square, repeated single factor experiments, etc. The data collection phase will occur during the experiments and will be automated due to the implementation of the Performance Evaluation Module within the expert system. Finally, data analysis will verify or reject the proposed hypotheses.

Since continued use of the expert system during experimentation increases task familiarity and expertise, the experimentation order influences the results. Thus, a Latin Square Design is appropriate for eliminating this effect. A Latin Square is a matrix that specifies different combinations of experimental

order for the same groups of subjects. For example, in an experiment with three levels of aiding (full, partial, and none), three groups of subjects would participate in different orders with the levels. (See Figure 9). Thus, group A would use full aiding first, partial aiding second, and no aiding third. However, group C would use no aiding first, full aiding second, and partial aiding third, etc. In addition to the Latin Square, other economical designs will be used according to Crolotte, A. and Dechter, R., "Plan for TCO Testing and Evaluation", Perceptronics Report PFTR-1082-11-79, November 15, 1979. These designs include one-factor two-level experiments such as randomized repetition, cross-over repetition, and non-rectangular Latin Squares, etc. (See Figure 10)

Data analysis for these designs consists of a one-way analysis of variance (ANOVA). A summary of the one-way ANOVA method is described in Figure 11.

<u>EXPERIMENT ORDER</u>	<u>NO AIDING</u>	<u>PARTIAL AIDING</u>	<u>FULL AIDING</u>
1	C	B	A
2	B	A	C
3	A	C	B

Figure 9. Latin Square Design

ONE FACTOR - 2 LEVEL EXPERIMENT

Number of Subjects per Experiment	Preferred and Alternative Designs	Total Number of Subjects Required	Number of Tests per Subject	Number of df per Error	Comments
1	10 cross-over	10	2	8	Very powerful design affording a high number of degrees of freedom for the error but fairly costly in number of subjects.
	7 cross-over	7	2	5	Powerful design requiring less subjects and still providing the minimum number of degrees of freedom for the error.
	6 randomized blocks	6	2	5	Design requiring less subjects than cross-over for the same number of degrees of freedom for the error. It is, however, implicitly assumed that no position effect exists.
2	7 cross-over	14	2	5	Same as above.
	6 randomized blocks	12	2	5	Same as above.
3	8 cross-over with repetition	12	4	4	If no treatment x cross-over effect exists, the number of degrees of freedom for the error is 5.
	4 randomized blocks with repetition	12	4	6	If no treatment x block effect exists, the number of degrees of freedom for the error is 7. No position effect assumed.
4	4 cross-over with repetition	16	4	4	Same as above.
	3 randomized blocks with repetition	12	4	4	If no treatment x block effect exists, the number of degrees of freedom for the error is 5. No position effect assumed.
≥ 5	3 randomized blocks with repetition	≥ 15	4	4	Same as above.

Figure 10. Contingency plans for one-factor 2-level experiments.

ONE-WAY ANALYSIS OF VARIANCE

One Way ANOVA

Source	Degrees of freedom (df)	Sum of Squares (SS)	Mean of Squares (MS)
Between Alternatives	$k-1$	$SS_{alt} = \sum_{j=1}^k \frac{T_{.j}^2}{n_j} - \frac{T^2}{N}$	$MS_{alt} = \frac{SS_{alt}}{k-1}$
Within Alternatives or Error	$N-k$	$\sum_{j=1}^k \sum_{i=1}^{n_j} y_{ij}^2 - \sum_{j=1}^k \frac{T_{.j}^2}{n_j}$	$MS_{err} = \frac{SS_{err}}{N-k}$
Totals	$N-1$	$\sum_{j=1}^k \sum_{i=1}^{n_j} y_{ij}^2 - \frac{T^2}{N}$	

Where y_{ij} = observation i of alternative j .

k = number of alternatives.

n_j = number of observations for alternative j .

$$N = \sum_j n_j$$

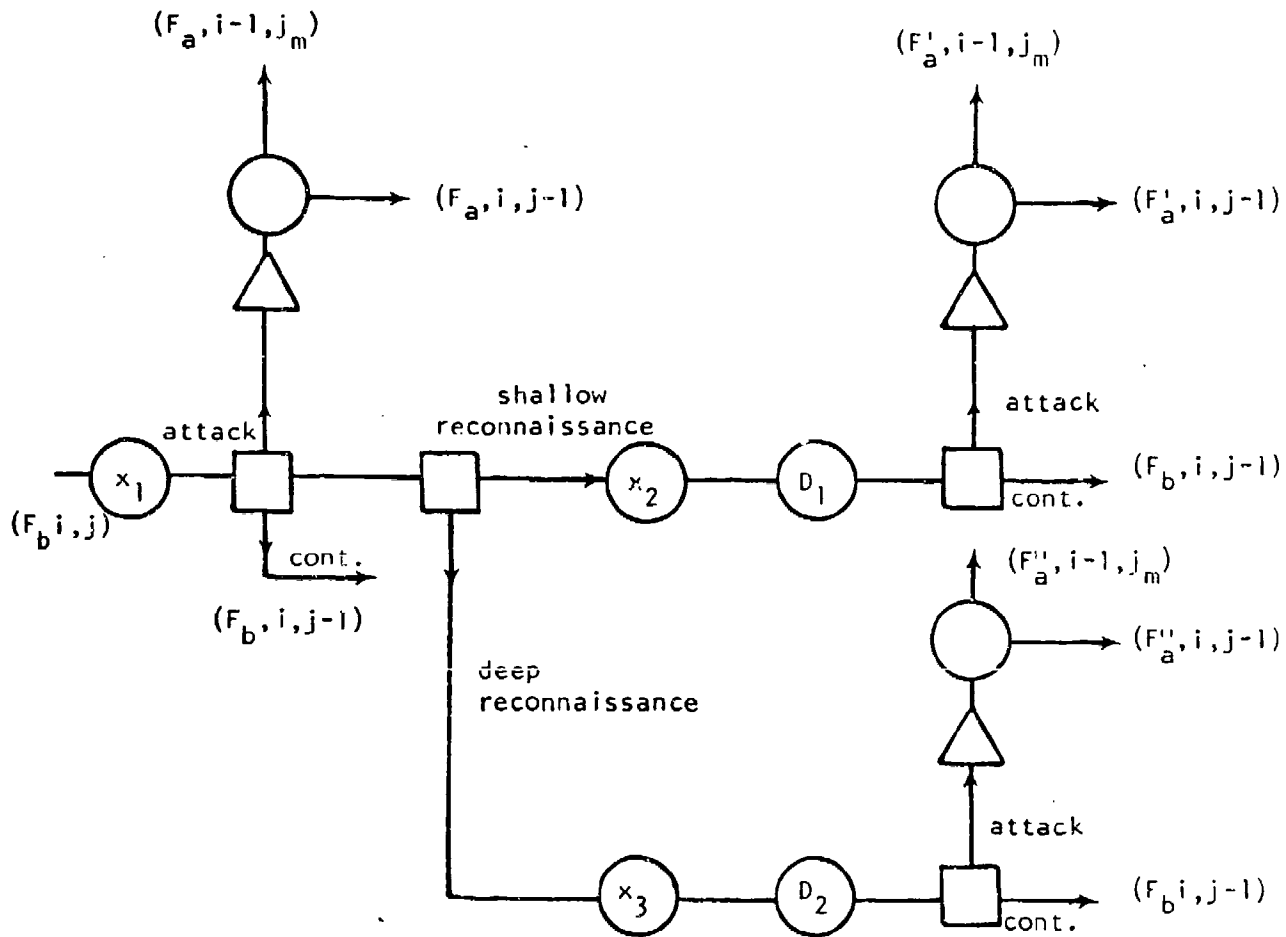
$$T_{.j} = \sum_{i=1}^{n_j} y_{ij} = \text{total for alternative } j.$$

T = Grand total.

Figure 11. One-way analysis of variance.

APPENDIX A DECISION DIAGRAM FOR GAME

The following decision diagram shows the decision points, event points, report points, and identification points for the Scenario Generator game.



Flow-diagram for data and decisions in the (i,j) stage. The tripiet (F,i,j) denotes the state of entering the (i,j) stage with force F .

Random Variables

F_b - Own force on entry

F_a - Own force after an attack

x_1 - Initial report regarding enemy strength

x_2 - Report obtained by shallow reconnaissance

x_3 - Report obtained by deep reconnaissance

D_1 - Being detected during shallow reconnaissance

D_2 - Being detected during deep reconnaissance

APPENDIX B

DECISION TABLES FOR THE GAME

The following tables show the optimal decisions for all encountered situations in the Scenario Generator game.

(I=4, J=7)

B-2

OPTIMALITY March 6, 1982

[illegible]

DECISION MAP

UNCLASSIFIED
(REPORT X1)

DECLASSIFIED (RE)

(REPORT X2)

NOT DETECTED

•

2731

FH-H	FMCA
HMM	GACA
THH	GCCA
HMH	GCCA
MHM	GCCA
LHM	GCCA
LMM	GCCA
NHM	GCCC
NMM	GCCC
NNL	GCCC

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

[illegible]

OPTIMALITY March 6, 1982

[illegible]

[illegible]

[illegible]

OPTIMALITY March 6, 1982

[illegible]

(I=2, J=6)

ENCAMPMENT 2

ENEMY LINE 3

DECISION MAP

LOW/DEEP (REPORT X1)

(REPORT X1)

[illegible]

US33333333
(REPORT X2)

[illegible]

(REPORT X2)

[illegible]

..ATTACK/CON

[illegible]

(REPORT X3)

[illegible]

FBI

$(i=2, j=5)$

ENCAMPMENT 3

ENEMY LINE 3

[illegible]

DECISION MAP

ATTACK/CONTINUE/SHALLOW/DEEP (REPORT X1)

[illegible][illegible]

ATTACK/CONTINUE	GIVEN SHALLOW, DETECTED	(REPORT X2)	CCCCCCCCCCCCCCCCCCCC
HHH	AAAAAIAACACCAAAAAA	CCCCAA	CCCCCCCCCCCCCCCCCCCC
HHM	AAAAAIAACACCCACACC	CCCCCC	CCCCCCCCCCCCCCCCCCCC
MHH	AAAAAIAACACCCACACC	CCCCCC	CCCCCCCCCCCCCCCCCCCC
MHM	AAAAAIAACACCCACACC	CCCCCC	CCCCCCCCCCCCCCCCCCCC
MHM	AAAAAIAACACCCACACC	CCCCCC	CCCCCCCCCCCCCCCCCCCC
MHM	AAAAAIAACACCCACACC	CCCCCC	CCCCCCCCCCCCCCCCCCCC
MHM	AAAAAIAACACCCACACC	CCCCCC	CCCCCCCCCCCCCCCCCCCC
LHM	AAAAAIAACACCCACACC	CCCCCC	CCCCCCCCCCCCCCCCCCCC
LHM	ACACCCACACCCCCCCC	CCCCCC	CCCCCCCCCCCCCCCCCCCC
NHM	ACACCCACACCCCCCCC	CCCCCC	CCCCCCCCCCCCCCCCCCCC
NHM	ACACCCACACCCCCCCC	CCCCCC	CCCCCCCCCCCCCCCCCCCC
NLL	CCCCCCCCCCCCCCCC	CCCCCC	CCCCCCCCCCCCCCCCCCCC
ATTACK/CONTINUE	GIVEN DEEP, NOT DETECTED	ATTACK/CONTINUE	GIVEN DEEP, DETECTED

HMLN		HMLN
HMLL		HMLL
HMLN		HMLN
CCAA	HHH	CCAA
CCCA	HHM	CCCA
CCCA	HHH	CCCA
CCCA	HHM	CCCA
CCCA	MMH	CCCA
CCCA	MMM	CCCA
CCCA	LHM	CCCA
CCCA	NHM	CCCA
CCCA	NNM	CCCA
CCCC	NLL	CCCC

(REPORT X3)

HMLN		HMLN
HMLL		HMLL
HMLN		HMLN
CCAA	HHH	CCAA
CCAA	HHM	CCAA
CCAA	HHH	CCAA
CCAA	HHM	CCAA
CCCA	MMH	CCCA
CCCA	MMM	CCCA
CCCA	LHM	CCCA
CCAA	NHM	CCAA
CCAA	NNM	CCAA
CCCA	NLL	CCCA

ENCAMPMENT 5

ENERGY LINE 3

[illegible]

OPTIMALITY March 6, 1982

DECISION MAP	ENEMY LINE 3	ENCAMPMENT 7	(I=2, J=1)
ATTACK/CONTINUE/SHALLOW/DEEP	HHH HHM HHM HHM HHM HHM LHM LHM LHM NMM NLL	HHH HHM HHM HHM HHM HHM LHM LHM LHM NMM NLL	HHH HHM HHM HHM HHM HHM LHM LHM LHM NMM NLL
ATTACK/CONTINUE GIVEN SHALLOW, NOT DETECTED	HHH HHM HHM HHM HHM HHM LHM LHM LHM NMM NLL	HHH HHM HHM HHM HHM HHM LHM LHM LHM NMM NLL	HHH HHM HHM HHM HHM HHM LHM LHM LHM NMM NLL
ATTACK/CONTINUE GIVEN SHALLOW, DETECTED	HHH HHM HHM HHM HHM HHM LHM LHM LHM NMM NLL	HHH HHM HHM HHM HHM HHM LHM LHM LHM NMM NLL	HHH HHM HHM HHM HHM HHM LHM LHM LHM NMM NLL
ATTACK/CONTINUE GIVEN DEEP, NOT DETECTED	HHH HHM HHM HHM HHM HHM LHM LHM LHM NMM NLL	HHH HHM HHM HHM HHM HHM LHM LHM LHM NMM NLL	HHH HHM HHM HHM HHM HHM LHM LHM LHM NMM NLL
ATTACK/CONTINUE GIVEN DEEP, DETECTED	HHH HHM HHM HHM HHM HHM LHM LHM LHM NMM NLL	HHH HHM HHM HHM HHM HHM LHM LHM LHM NMM NLL	HHH HHM HHM HHM HHM HHM LHM LHM LHM NMM NLL

[illegible]

OPTIMALITY MATCH 6, 1932

DECISION MAP	ENEMY LINE 4	ENCAMPMENT 2	(I=1, J=6)
ATTACK/CONTINUE/SHALLOW/DEEP	<p> HHH HHM HHM HHM HHM HHM LHM LHM NMM NMM NLL </p>	<p> HHH HHM HHM HHM HHM HHM LHM LHM NMM NMM NLL </p>	<p> HHH HHM HHM HHM HHM HHM LHM LHM NMM NMM NLL </p>
ATTACK/CONTINUE GIVEN SHALLOW, NOT DETECTED (REPORT X1)	<p> HHH HHM HHM HHM HHM HHM LHM LHM NMM NMM NLL </p>	<p> HHH HHM HHM HHM HHM HHM LHM LHM NMM NMM NLL </p>	<p> HHH HHM HHM HHM HHM HHM LHM LHM NMM NMM NLL </p>
ATTACK/CONTINUE GIVEN SHALLOW, DETECTED (REPORT X2)	<p> HHH HHM HHM HHM HHM HHM LHM LHM NMM NMM NLL </p>	<p> HHH HHM HHM HHM HHM HHM LHM LHM NMM NMM NLL </p>	<p> HHH HHM HHM HHM HHM HHM LHM LHM NMM NMM NLL </p>
ATTACK/CONTINUE GIVEN DEEP, NOT DETECTED (REPORT X3)	<p> HHH HHM HHM HHM HHM HHM LHM LHM NMM NMM NLL </p>	<p> HHH HHM HHM HHM HHM HHM LHM LHM NMM NMM NLL </p>	<p> HHH HHM HHM HHM HHM HHM LHM LHM NMM NMM NLL </p>
ATTACK/CONTINUE GIVEN DEEP, DETECTED	<p> HHH HHM HHM HHM HHM HHM LHM LHM NMM NMM NLL </p>	<p> HHH HHM HHM HHM HHM HHM LHM LHM NMM NMM NLL </p>	<p> HHH HHM HHM HHM HHM HHM LHM LHM NMM NMM NLL </p>

ENGAGEMENT 3

ENERGY LINE 4

DECISION MAP

LOW/DEEP (REPORT XI)

(REPORT X1.)

ATTACK/CONTINUE/SHALLOW/DEEP

[illegible]

REPORT X2)

1122
ATTACK/CONTINUE GIVEN SHALLOW, NOT DETECTED

[illegible]

(REPORT X2)

ATTACK/CONTINUE GIVEN SHALLOW, DETECTED

[illegible]

...ATTACK/CONF

NEE
ATTACK/CONTINUE GIVEN DEEP, NOT DETECTED

HMNL HMNL
HMLL HMLL
HMNL HMNL
(REPORT X3)
HMNL

(REPORT X3)

H H H H H H H H H H

HHH
HHH
HHH
HHH
HHH
HHH
LHM
LHM
NMH
NMH
NLL

[illegible]

DECISION MAP

REPORT X1)

ATTACK/CONTINUE/SHALLOW/DEEP

ASST X21 REPORT

ATTACK/CONTINUE GIVEN SHALLOW, NOT DETECTED

REPORT

**NEC
ATTACK/CONTINUE GIVEN SHALLOW, DETECTED**

ATTACK/CON

NEE
ATTACK/CONTINUE GIVEN DEEP, NOT DETECTED

2013

71-15418-4

117344

U.S. 11

CAA: A

CAA
CAA

LA 13
LA 13

U
U
A
A
A
A

CAA:3

CAA:A

ALL

五五

550

DECISION MAP

ENEMY LINE 4

ENCAMPMENT 6

(I=1, J=2)

[illegible][illegible][illegible][illegible][illegible]

$(I=1, J=1)$

ENCAPMENT 7

ENEMY LINE 4

NNJNLLNNJNLLNNLLMMNNHLLMMNNMNNLLHHNNMMHHLLHLMMNFHLLMMHHMLMMHHMMHH
 NNJNLLNNJNLLMMNNLNNHLLMMFLMMFLHHNNLMMHNNMMLHLMLMMHMMHLLHLMMHLMHMMHH
 NLNNLMMNNLMLMNLNNHNNMLMMNNHLLHNLNMMHNNMNFLLMMLHHNNHMLMLMMHLLHMMHHMMH
 ATTACK/CONTINUE/SHALLOW/DEEP (REPORT X1)

```

HHH AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
HHM AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
HHH AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
HHM AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
HHM AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
HHM AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
MMM AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
LHM AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
LHM AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
NMM AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
NMM AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
NLL AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
ATTACK/CONTINUE GIVEN SHALLOW, NOT DETECTED (REPORT Y2)

```

[illegible][illegible]

HHH	HMLN		(REPORT X3)	FIH	
HHM	HMLL			FHM	
MHH	HMLN			MHH	
AAA	AAAA			HMM	
AAA	AAAA			NHM	
AAA	AAAA			LHM	
AAA	AAAA			LHM	
AAA	AAAA			NHM	
AAA	AAAA			NHM	
AAA	AAAA			NLL	